A Bioinformatics Approach to Identify Microarray Gene Expression Toxicant Signature Patterns

Pierre R. Bushel[1], Hisham Hamadeh[1], Lee Bennett[1], Stella Sieber[1], Karla Martin[1], Kerry Blanchard[2], Supriya Jayadev[2], Olimpia DiSorbo[2], Ray Stoll[2], Cynthia A. Afshari[1], and Richard Paules[1]

[1]*NIEHS Microarray Center,* [2]*Boehringer Ingelheim Pharmaceuticals*

http://dir.niehs.nih.gov/microarray

bushel@niehs.nih.gov

Cambridge Healthtech Institute

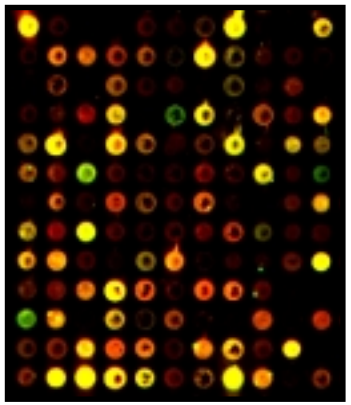
Bioinformatics and Genome Research

June 19-20, 2000

The Fairmont Hotel, San Francisco, California

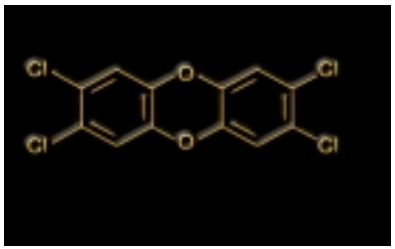
Toxicogenomics

Microarray
Gene Expression Data



+

Toxicological Information



Toxicogenomics Database

Heuristic Knowledge

Toxicant Hazard Identification, Characterization and Risk Assessment

Similar database model in the form of a pharmacological database, Scherf, et al. Nature Genetics 24, March 2000

Abstract

DNA microarray technology has emerged to the forefront of gene expression analysis as a tool by which researchers can detect genome-wide differential expression of thousands of genes. Recent interest in identifying transcription signature patterns has led to an increase in the use of microarray technology to simultaneously analyze, monitor and characterize changes in gene expression profiles in response to serum induction, cell cycle changes, cellular processes, genotoxic stress and oncogenesis.

The National Institute of Environmental Health Sciences Microarray Center (NMC) is combining the fields of toxicology and genomics to better understand mechanistic based risk assessment, predictive toxicology and hazard identification of a variety of compounds. This new sub-discipline, termed toxicogenomics, stems from the use of high-density microarray technology and toxicology to measure changes in gene expression patterns that are different in biological models following exposure to toxic agents. Essential to this effort are bioinformatics, computational biology and statistical analysis which lend the biological informatics resources, robust computing power and mathematical methodologies to confidently correlate gene expression profiles of unknown agents with the signature patterns of known toxicants to ultimately link gene expression information with toxicological endpoints.

In order to discern whether a specific signature pattern for a class of compounds could be elucidated, microarray gene expression data were generated from analysis of liver samples from Sprague-Dawley rats following treatments by three “classic” peroxisome proliferator compounds and contrasted with treatment with phenobarbital, a barbituate with a different mode of action, and D-mannitol, a negative control. The results were subjected to various forms of data analysis to extract meaningful gene expression pattern relationships. ArraySuite statistical analysis software selects outlier genes based on confidence intervals computed from the distribution of all ratio outlier values. Candidate outlier genes are statistically validated in the NMC MicroArray Project System (MAPS) by using a multinomial distribution to determine the probability of identifying random outliers in replicate experiments. An overview of compiled validated outlier genes reveals a general differential gene expression relationship between the peroxisome proliferators when data points are ranked according to the Wyeth 14,643 treatment ratio outlier values. Pair-wise comparisons of differentially expressed genes clearly demonstrate strong correlation ($R > +0.8$) between independent replicate biological samples, good correlation ($R > +0.5$) between peroxisome proliferators and little correlation ($R < +0.4$) between the peroxisome proliferators and phenobarbital. Finally, two-dimensional hierarchical cluster analysis partitions genes with associated mechanistic pathways into highly correlated nodes ($R > +0.8$) that are indicative of specific differential signature patterns that are associated with peroxisome proliferator treatments.

Methodology

We investigated the gene expression profile of various toxicants using microarray technology with a rat chip containing ~1700 cDNA clones. Details of the protocols and analysis software used can be found on the NIEHS Microarray Center (NMC) web site at http://dir.niehs.nih.gov/microarray.

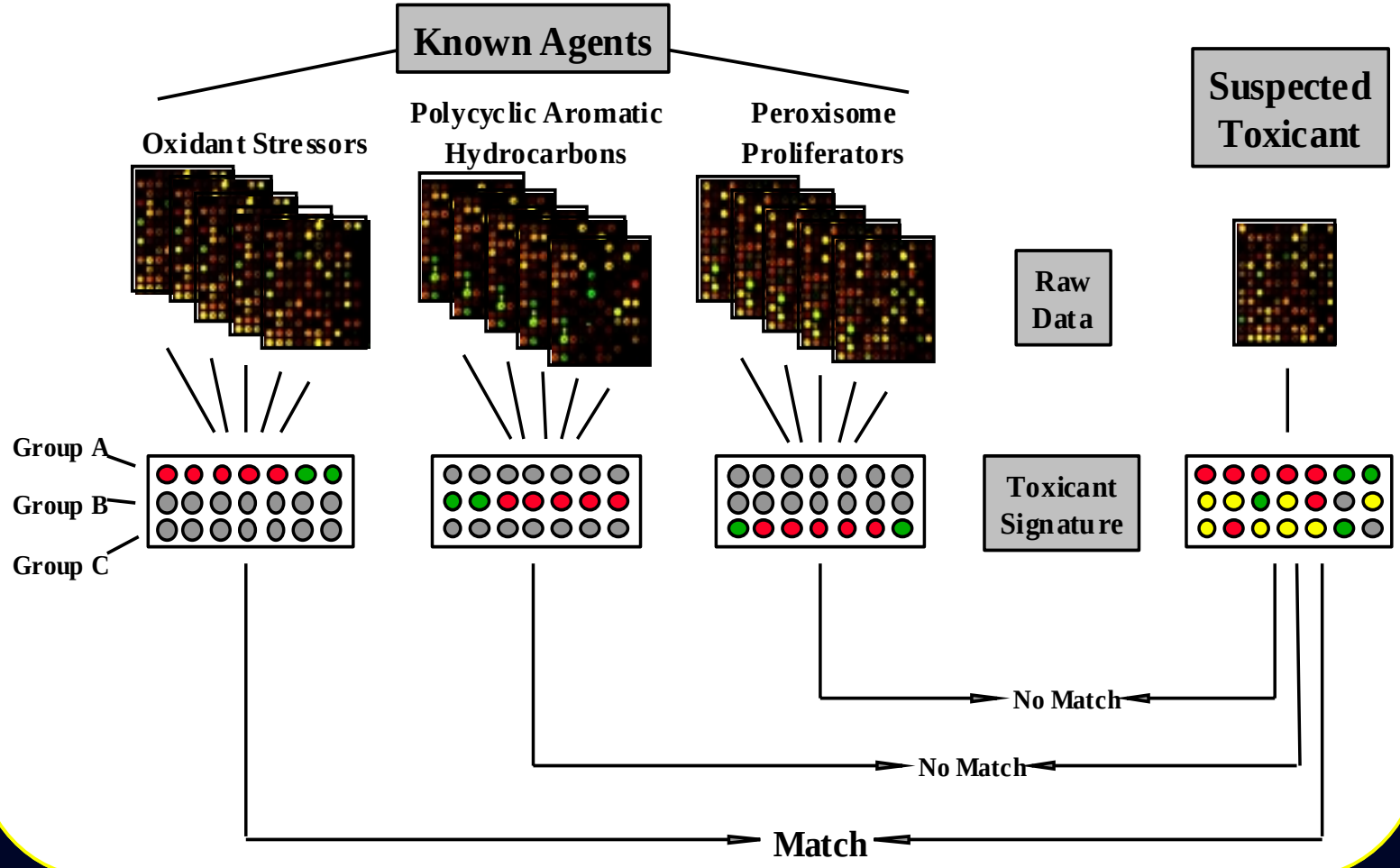
Sprague Dawley male rats (3 per compound) were exposed to toxicants for 24hrs. or 2 weeks. Doses of each compound for both time points were as follows:

Clofibrate	250 mg/Kg/day
Wyeth 14,643	250 mg/Kg/day
Gemfibrozil	100 mg/Kg/day
Phenobarbital	120 mg/kg/day
D-mannitol	500 mg/Kg/day

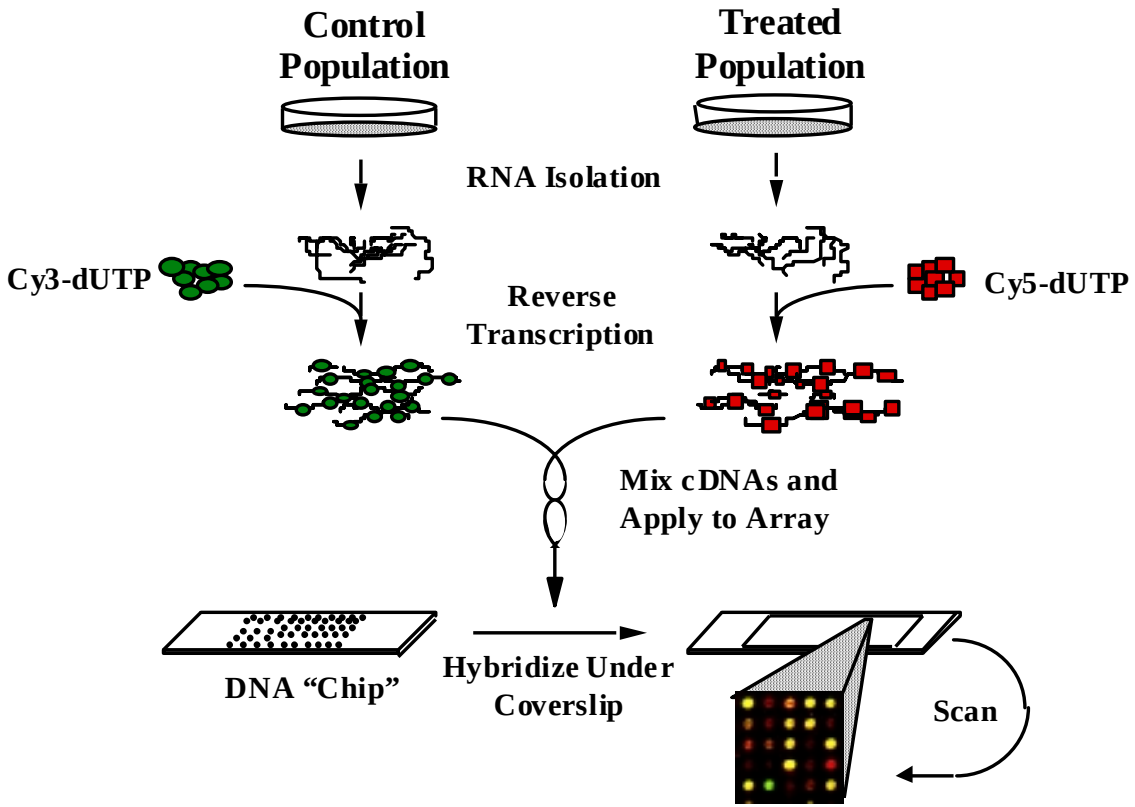
RNA Samples were prepared from extracted livers of animals exposed to the compounds and labeled with fluorescent molecules. cDNAs made from each RNA sample were hybridized in triplicate to rat microarray chips. A pooled sample, prepared from livers of untreated male rats, was used as a control in all hybridizations.

Scanned images were processed and analyzed using IPLab software with ArraySuite extensions and data was archived in our internal MicroArray Project System (MAPS) database. Spotfire Pro, Cluster, and TreeView software were used to further analyze the gene expression data.

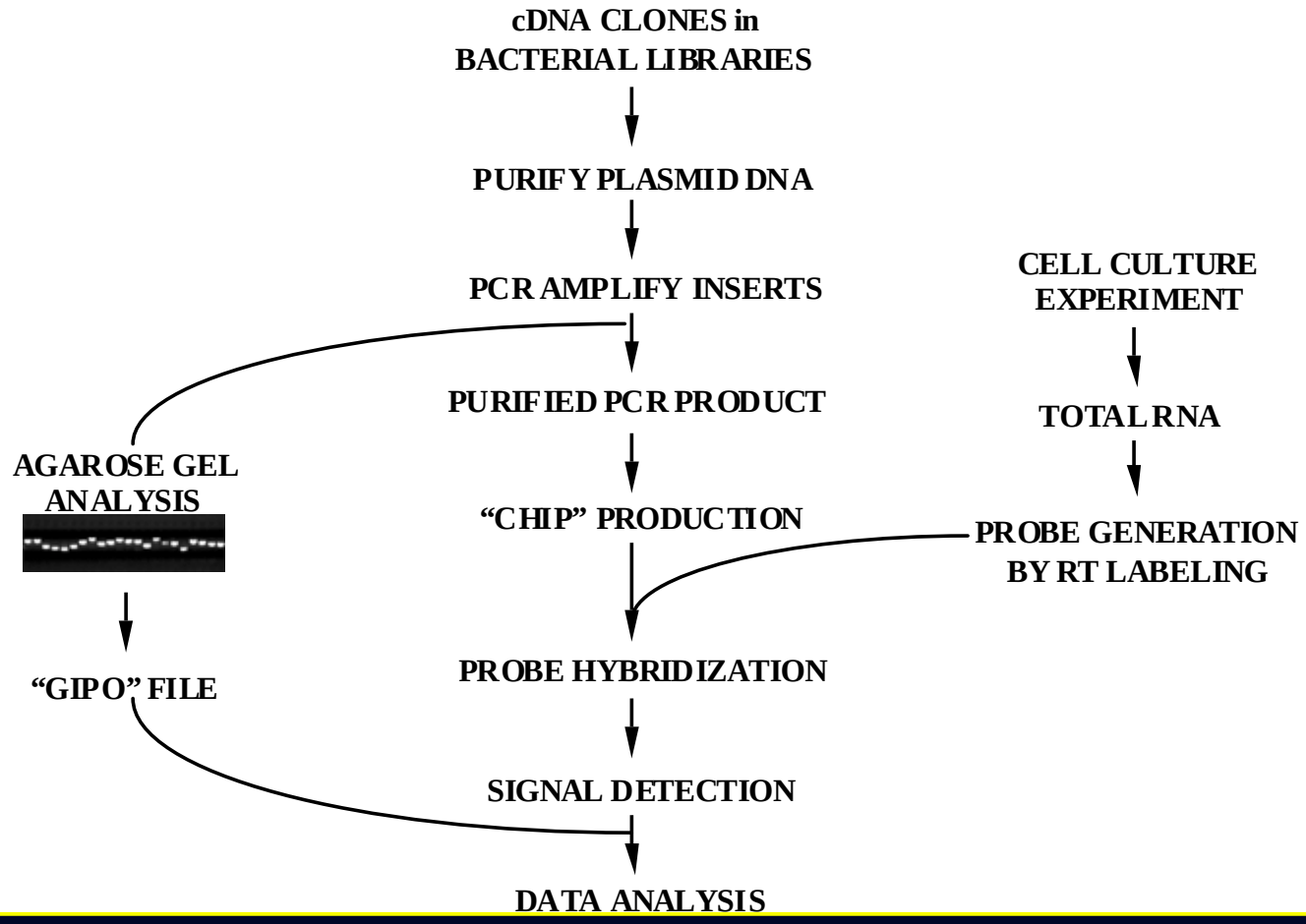
Toxicant Identification and Classification Using cDNA Microarrays



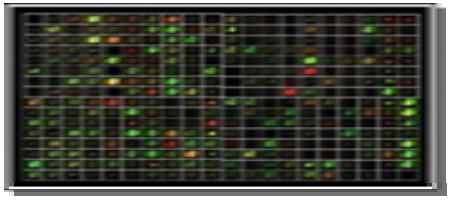
Simplified Overview of Gene Expression Analysis Using cDNA Microarrays



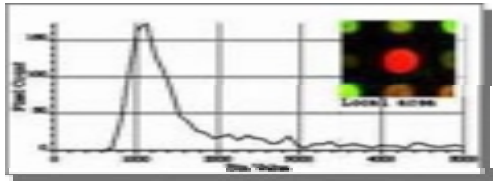
Procedure Overview



ARRAYSUITE SOFTWARE: Order of Operations

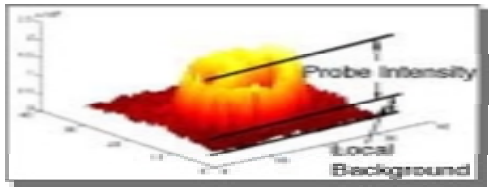


1. Target Segmentation

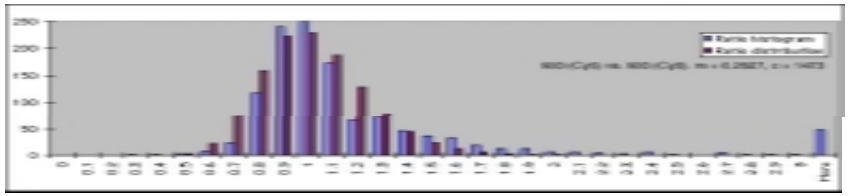


2. Background Subtraction

3. Target Detection



4. Target Intensity Determination



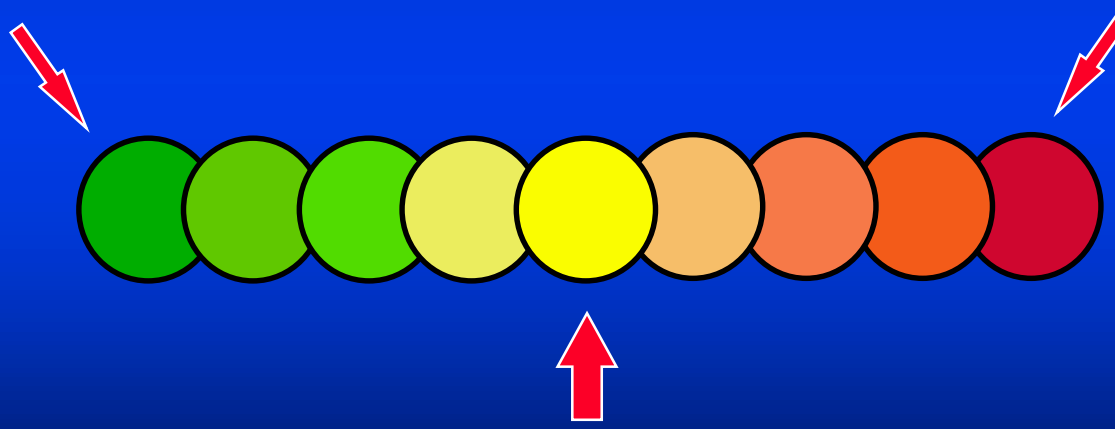
5. Ratio Analysis

Images courtesy of Dr. Y. Chen, NHGRI

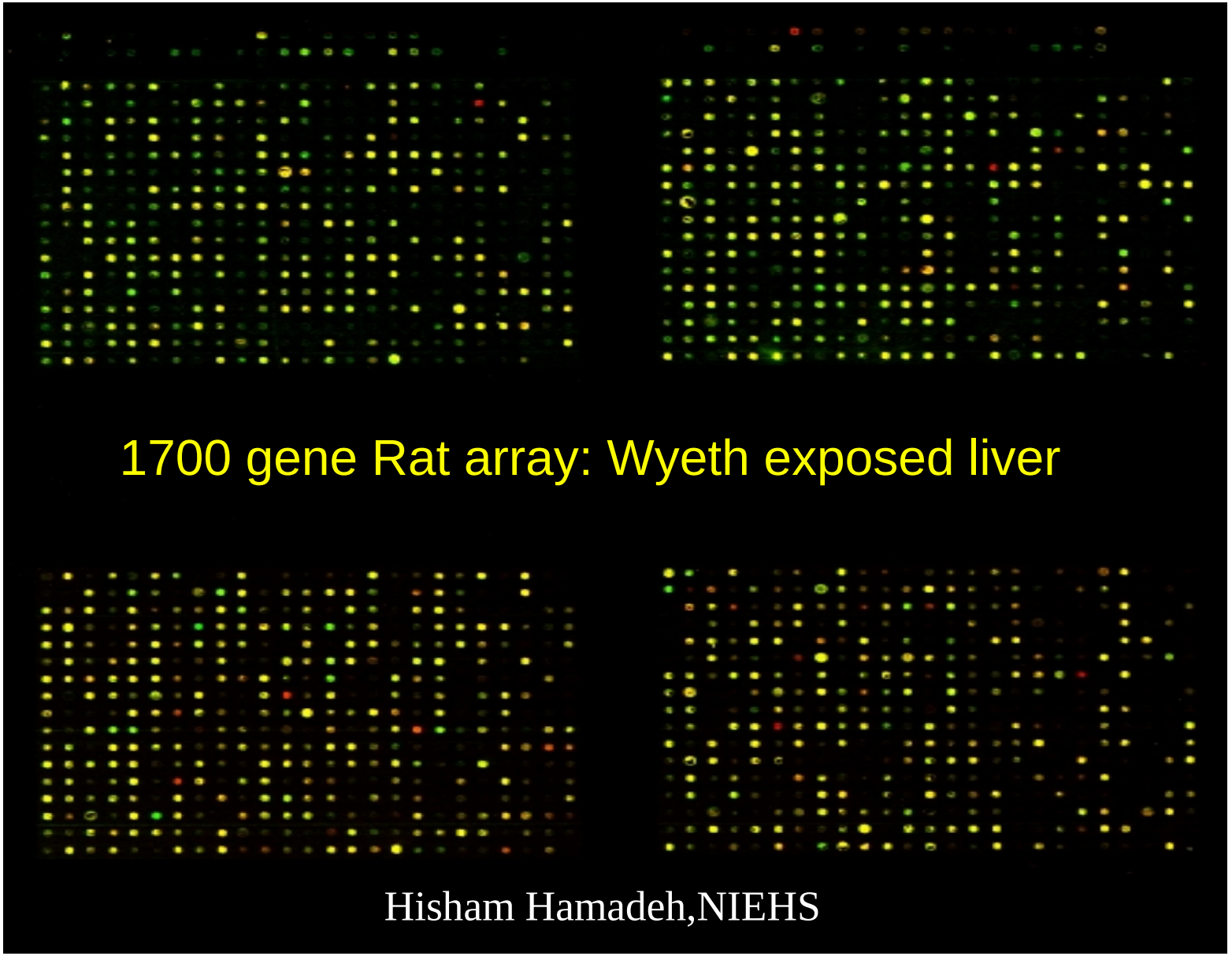
Data Collected from Two-Color Hybridizations

mRNA species present exclusively in population A

mRNA species present exclusively in population B



mRNA species present at equal levels in A and B


1700 gene Rat array: Wyeth exposed liver
Hisham Hamadeh,NIEHS

Statistical Analysis of Data Using ArraySuite

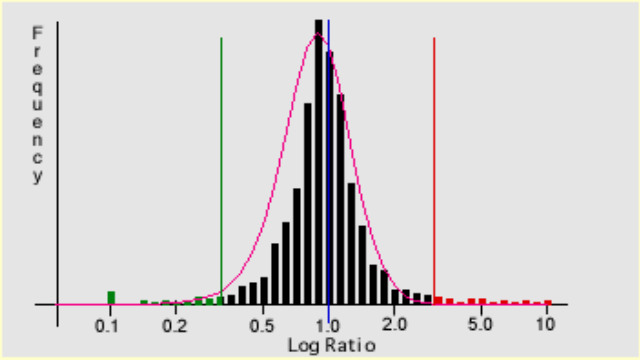
Image Name: p30s08c3bmarkrotflp
p30s08c5b.Tif

Histogram/ Scatter Plot
Ratio
Red channel
Green channel
Size

Data from:
All targets
Control targets

Calibration by:
Internal Controls
All targets
Background

Calibration Method:
Ratio Distribution
Log-Normal



Log Scale
Calibrated Result

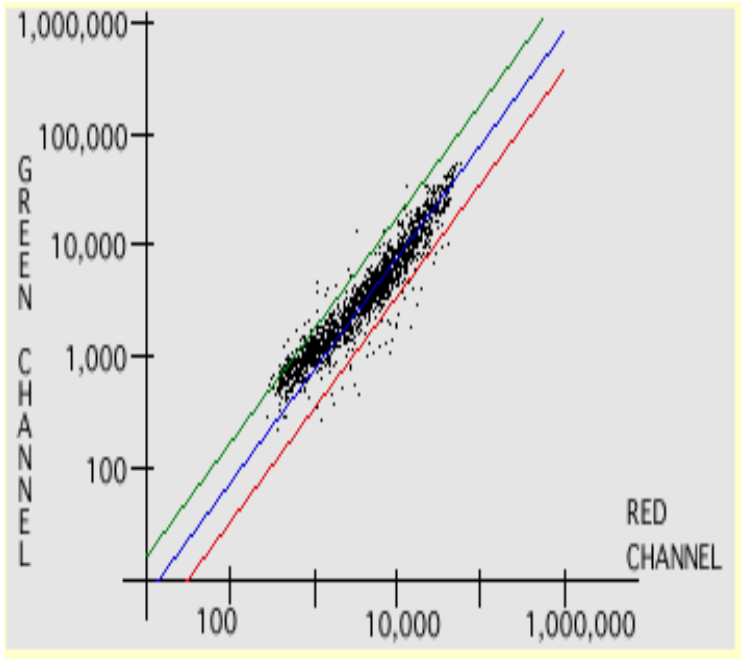
Confidence Level:	99.00	Intensity From:	1
Ratio Lower Limit:	0.32	Intensity To:	65535
Ratio Upper Limit:	3.10	Target Size From:	10

Refine Stats

CV = 0.250
M = 1.193

Default Setting | Ratio Outliers | Exit | Refresh
Update dataSheet | Equal Intensities

TargetLocator, 1.3, © 1997, NIH/NHGRI/LCG



Total outliers: 77 This page: 33-43

ID	Ratio	Intensities	Description
AI145215	3.59	576.1/ 134.6	"Rattus norvegicus mRNA for ETR-R3b protein, alternatively spliced isoform"
AI030720	3.49	314.6/ 75.6	"Rattus norvegicus casein kinase 1 alpha (CK1a) mRNA, complete cds"
AI136539	3.36	392.8/ 97.8	"Rat heme oxygenase-2 (HO2) mRNA, complete cds"
AI045067	3.30	695.9/ 176.7	"Rattus norvegicus water channel protein (AQP3) gene, complete cds"
AA956162	3.25	20294.1/5236.6	"Rattus norvegicus G protein coupled receptor kinase 5 mRNA, complete cds"
AI029744	3.22	483.1/ 125.9	"Rattus norvegicus brain hexokinase mRNA, complete cds"
AA899106	3.20	511.1/ 134.0	"Rat cyclin D2 (VIN1) mRNA, complete cds"
AA900276	0.31	124.9/ 333.6	"Rat zinc finger protein (RP8) mRNA, 3' end"
AA858569	0.30	82.4/ 226.7	"Rattus norvegicus neuronatin beta mRNA, complete cds"
AA819821	0.30	151.6/ 425.9	"Rattus norvegicus lactate dehydrogenase-B (LDH-B) mRNA, complete cds"
AA956846	0.29	106.4/ 304.1	"Rat microsomal aldehyde dehydrogenase mRNA, complete cds"

Save | Print | Exit | Previous | Next

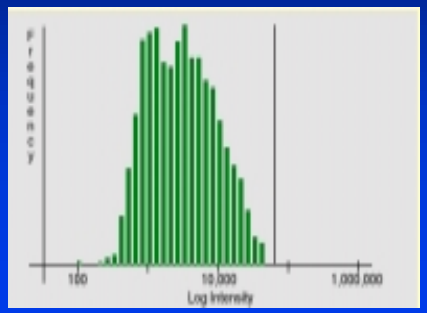
TargetLocator, 1.3, © 1997, NIH/NHGRI/LCG

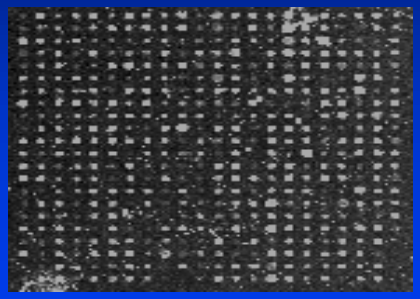
Target Intensity and Pixel Size

Intensity Channels

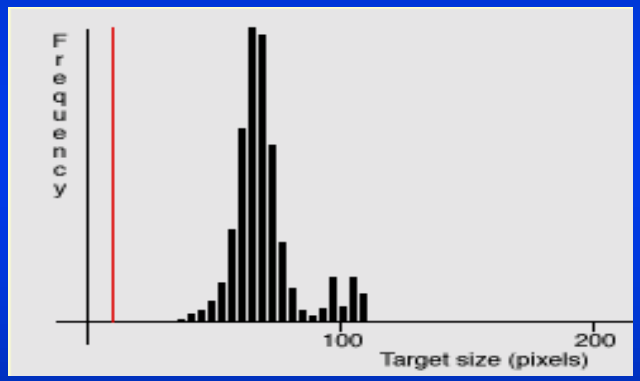
PMT=8.61

Cy5



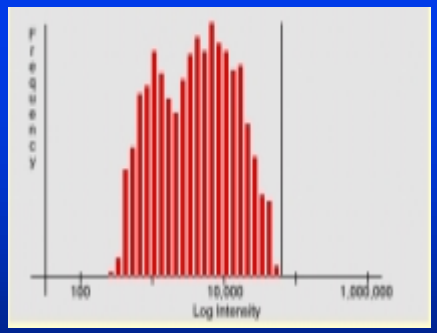


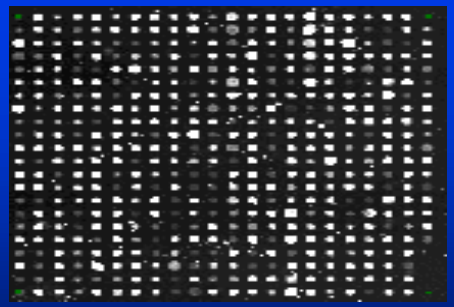
Target Size



PMT=4.6

Cy3





Probability of Random Outliers

1700 genes

Triplicate experiments (95% confidence)

Times flagged by chance	*Probability*	*Expected #*
0	0.85738	1482
1	0.13538	234
2	0.00713	12
3	0.00013	0

Quadruplicate experiments

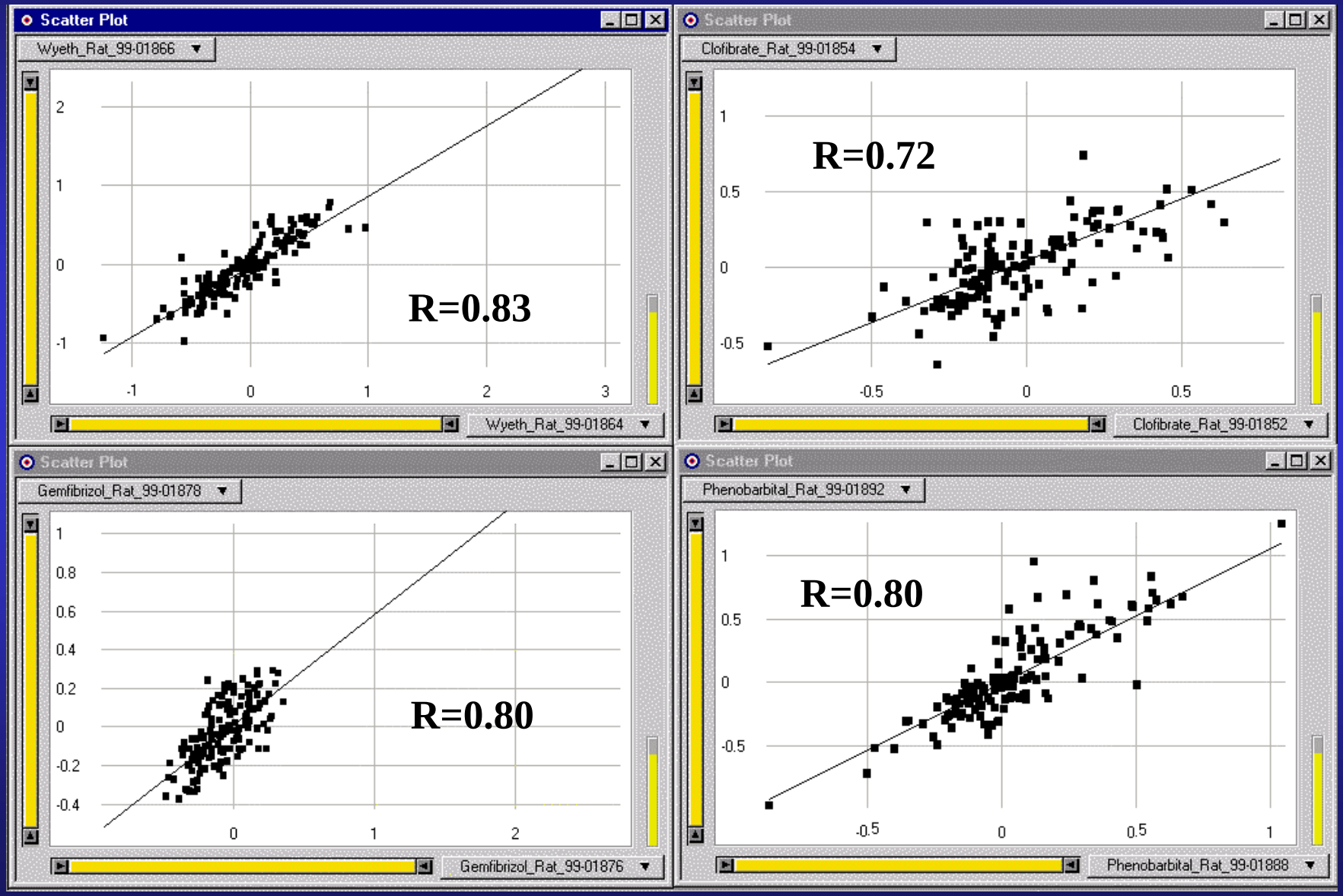
Times flagged by chance	*Probability*	*Expected #*
0	0.81451	1407
1	0.17148	296
2	0.01354	23
3	0.00048	1
4	0.00001	0

Probabilities computed using a binomial distribution

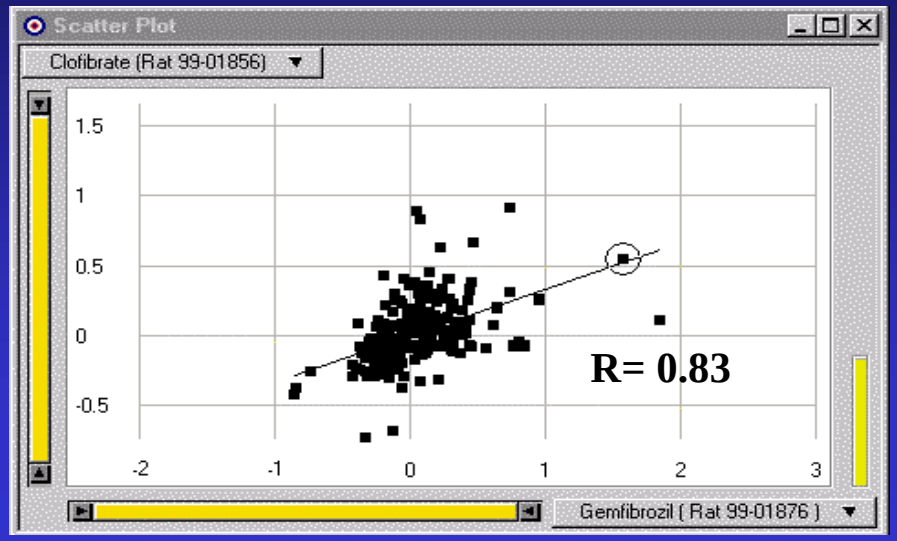
$$P(Y=y|n,p) = \binom{n}{y} p^{y}(1-p)^{n-y} \quad y=0,1,2,\ldots,n$$

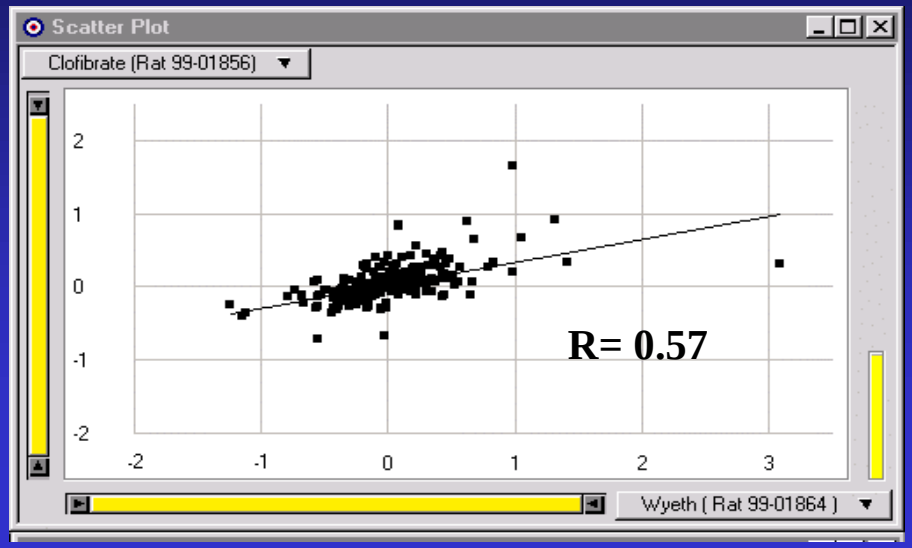
Lee Bennett, NIEHS

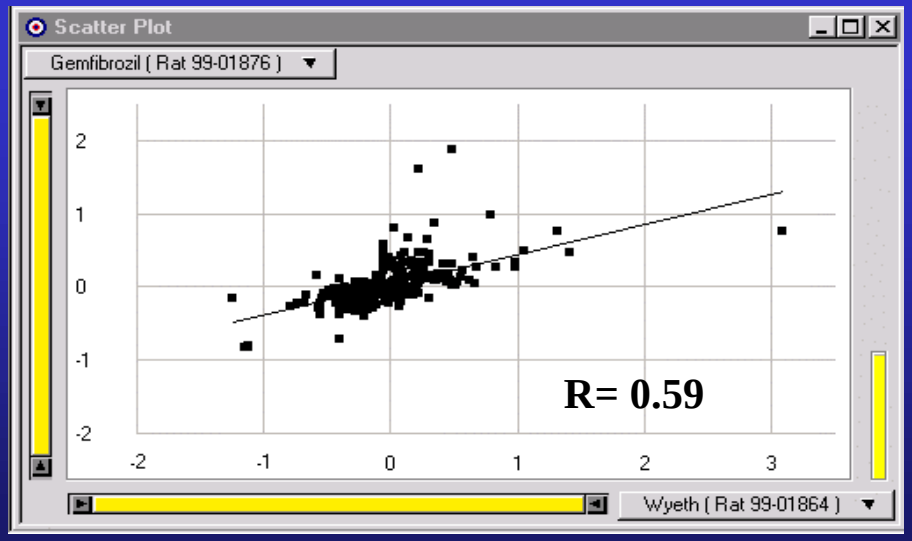
Correlation Between Animals

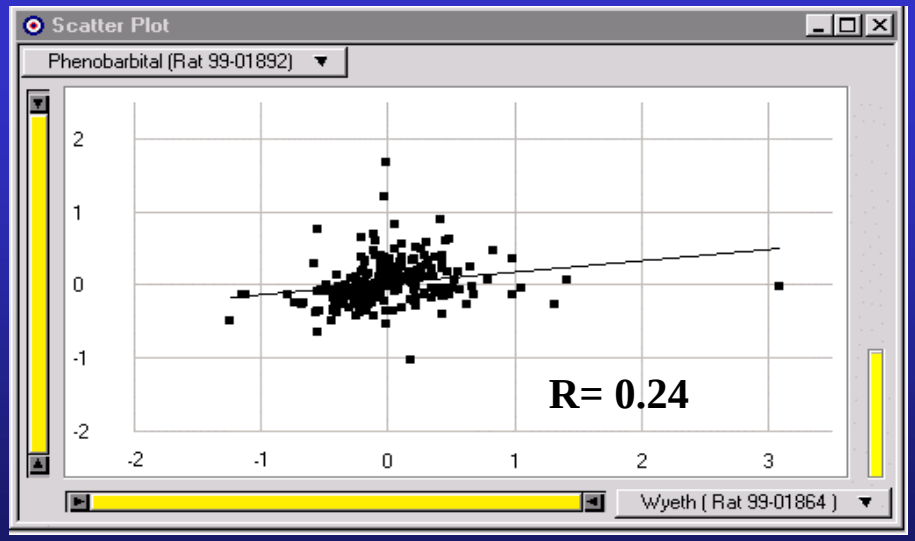


Correlation Between Compound Treatments

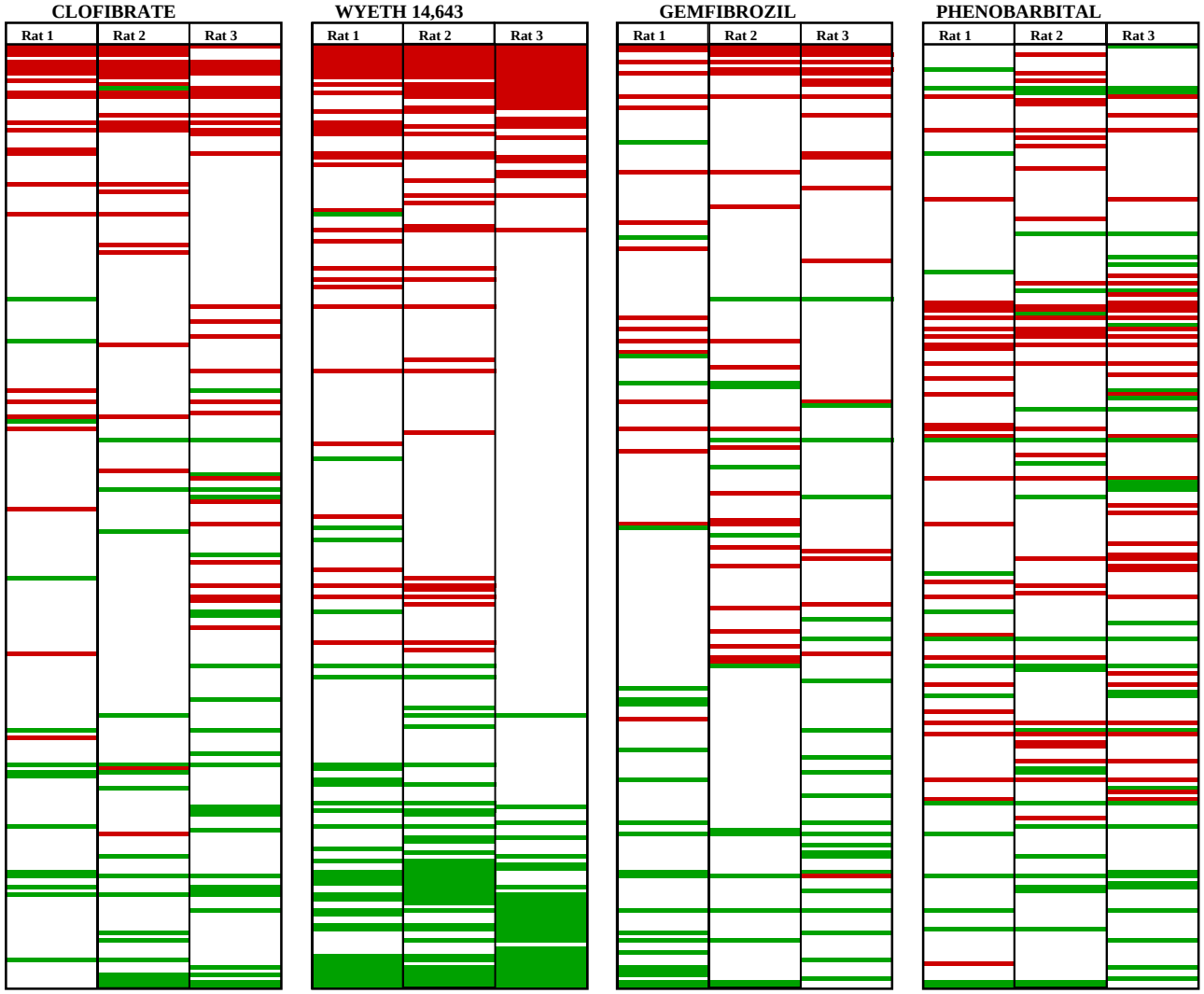




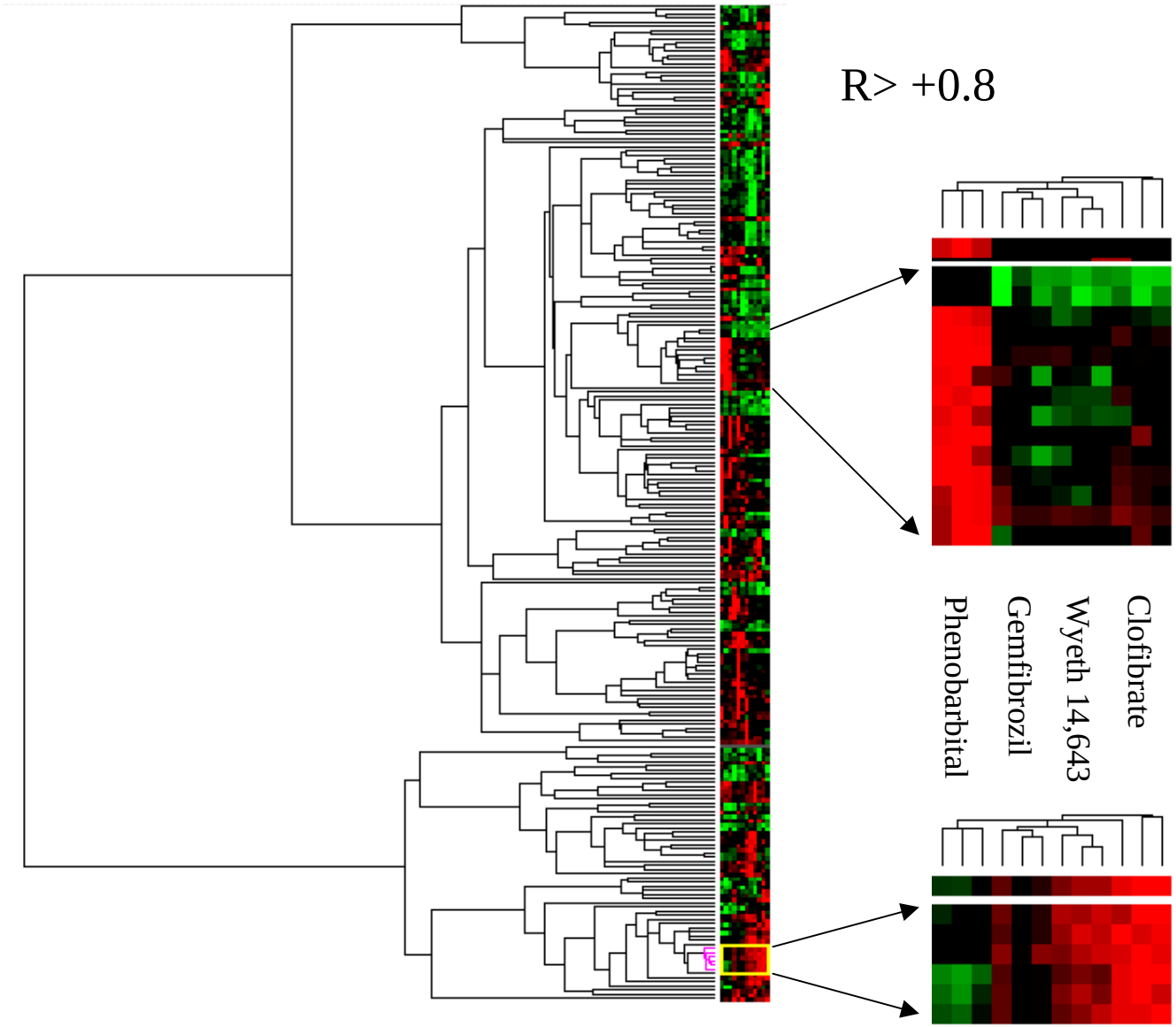




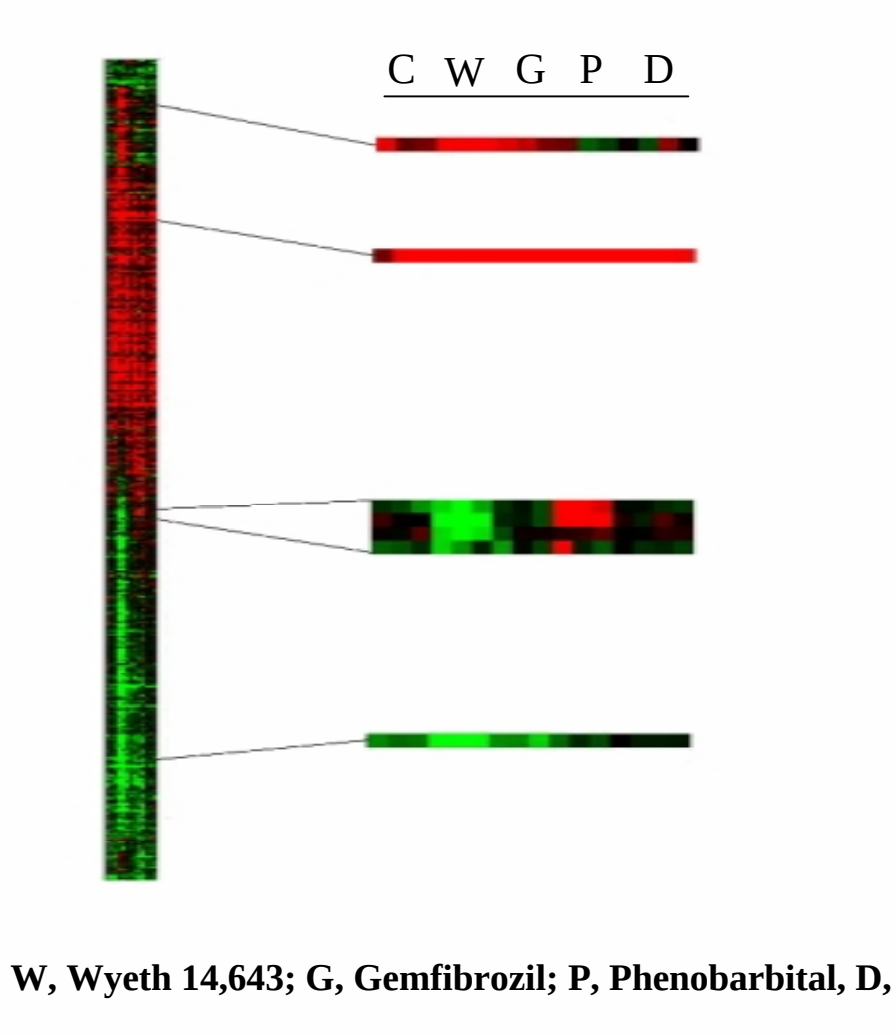
Finger Print "Bar Code" Comparison of Ratio Outliers



Hierarchical Cluster Analysis of Ratio Outliers from the24hr Treatment

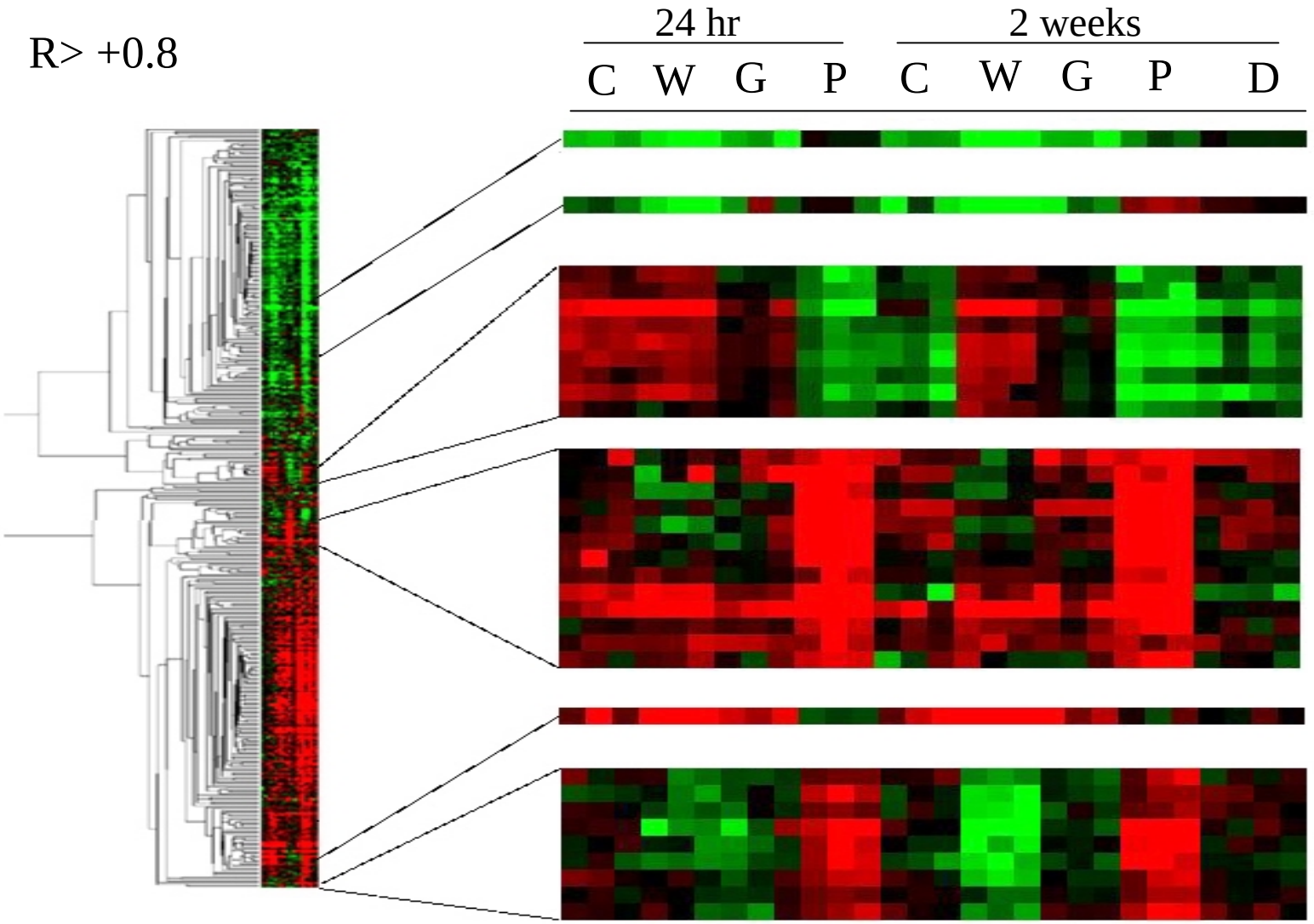


Hierarchical Cluster Analysis of Ratio Outliers from the Two-Week Treatment



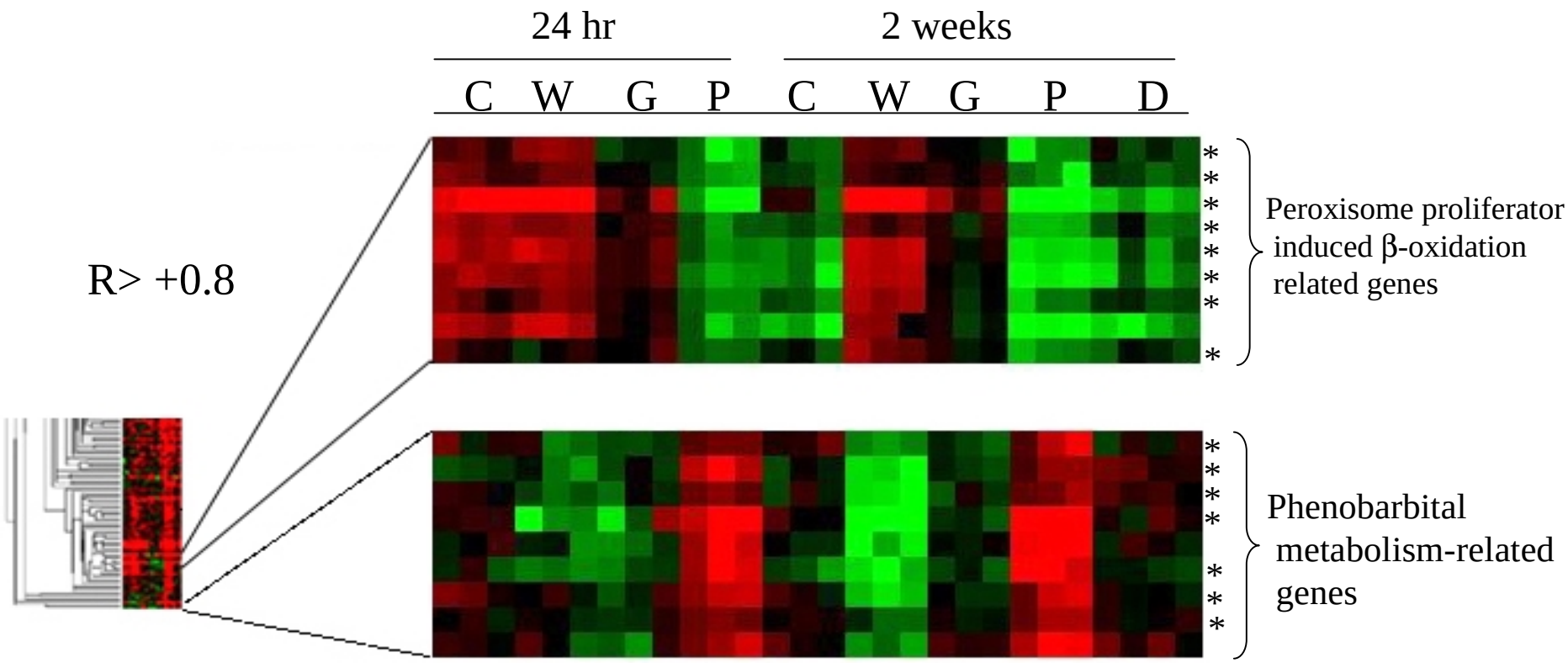
C, Clofibrate; W, Wyeth 14,643; G, Gemfibrozil; P, Phenobarbital, D, D-mannitol

Hierarchical Cluster Analysis of Ratio Outliers from 24hr and Two-Week Treatments



C, Clofibrate; W, Wyeth 14,643; G, Gemfibrozil; P, Phenobarbital, D, D-mannitol

Cluster Nodes of Genes with Similar Biological Function



C, Clofibrate; W, Wyeth 14,643; G, Gemfibrozil; P, Phenobarbital, D, D-mannitol

Conclusions

Using microarray technology and toxicology we distinguished between different classes of compounds such as peroxisome proliferators and phenobarbital through differences in their respective gene expression profiles.

Using bioinformatics tools to pair-wise compare and cluster gene expression data we observed highly correlated subsets of genes that are consistently differentially expressed at the two time points for the same compound. The genes may be part of a time-dependent gene expression profile associated with the respective compound.

Our genomics data is in agreement with the histopathology observed in the treated animals. We observed the up-regulation of structural genes such as tubulin in samples where liver enlargement was reported in the animal.

Using bioinformatics and genomics we better understand the feasibility of predicting the potential toxicity of an unknown compound under the working hypothesis outlined.

Work in Progress

The NIEHS Microarray Center and Boehringer Ingelheim Pharmaceuticals are in the initial stages of treating additional rat animal models as well as other organisms with compounds at multiple doses and time points for further detailed analysis of toxicant induced differential changes in microarray gene expression. It is also the interest of the collaboration to investigate toxicant signature patterns of genes in other tissues and organs and to use bioinformatics and proteomics to determine protein level changes in response to toxicant treatments.

The NIEHS Microarray Center is currently implementing an Oracle version of publicly available ArrayDB to provide users web based access to analyze microarray images and integrate processed gene expression data with internal and external biological resources. There is also a vast amount of interest in developing a toxicogenomics database as a central repository for storage and access to toxicological microarray gene expression data to gain heuristic knowledge for toxicant hazard identification, characterization, and predictive risk assessment.

Plans are currently underway to obtain and develop other bioinformatics, statistical, and computational biology software and methodologies such as principal component analysis (PCA), support vector machines (SVM), self organizing maps (SOM), genetic networks, reverse engineering and hierarchical clustering for higher-order analysis of multi-variate microarray gene expression data.